Spectrogram warping distance.

Matthew Sottile

November 15, 2005

1 Description of Metric

This metric is based on computing the similarity of short time Fourier representations of the signal at well defined windows. It also provides a metric of similarity for the evolution of a particular frequency component over these windows in time.

2 Mathematical principles

This metric compares two time series based on the changes in their frequency spectra over a moving short time window. The comparison takes into account how single frequency components vary over time, and differences in the spectrum at a specific time window between the series. Consider two time series, P and Q of length n_P and n_Q respectively, each representing discrete sampled values of a continuous signal sampled at T_s second intervals. The Nyquist sampling theorem dictates that the frequency components that can be perfectly reconstructed ranges from $\frac{1}{2T_s}$ Hz to $\frac{n}{2T_s}$ Hz. Often very low frequency information is of interest, followed by the behavior of intermediate range frequencies. High frequencies often represent noise. Furthermore, the presence of transient of non-stationary events should be detectable. As such, the short time Fourier transform can be employed to examine the spectrum of the signal over a short time window $w << n$. If one creates a sequence of spectra for a window that moves over the sequence, then the familiar spectrogram is created (eq. 1).

$$sg(X) = \left\{ \mathcal{F}(X_{[0,w]}), \mathcal{F}(X_{[w,2w]}), \mathcal{F}(X_{[2w,3w]}), \cdots, \mathcal{F}(X_{[kw,n]}) \right\} \tag{1}$$

Each element i of $sg(X)$ corresponds to the spectrum (a sequence of frequency components at $\omega_i = \frac{i}{2T_s}$, $1 \leq i \leq w$) over the ith window of w samples. Transient events can be easily identified as frequency components that are particularly strong in one or more elements of $sg(X)$ and weak in all others. Nonstationary events correspond to peaks in the spectrum that appear at different but neighboring frequencies as i increases. If one plots these spectra as a 2-dimensional image such that the x-axis corresponds to i, and the y-axis corresponds to frequency, then transients represent short horizontal lines, while non-stationary processes result in lines with either non-zero slope, or some form of non-linear curve.

The metric that this code produces is based on measuring the difference between two spectrograms due to mismatches in each window pair and each frequency component. This is implemented by computing the similarity metric for each column and each row in the 2D spectrogram. The sequence similarity is computed using the *dynamic time warping* algorithm.

3 Physical and Engineering Principles

The magnitude of frequency components at any given time give some indication of the rate of energy transfer into the system.

4 Usage

```
[Labels, Measures]=Metrics_Sottile(ref,data,params)
```

Input

`ref`	time series of reference data set
`data`	time series to compare with `ref`
`params`	A vector of parameters. `params(1)` represents the window size for the FFT.

Output

`Labels`	N by 1 cell array of strings corresponding to `Measures` array
`Measures`	N by 1 cell array of Measures labelled in `Labels`

Notes

The current implementation returns three measures ($N = 3$).

`Measures{1}`	Total DTW warp cost in frequencies
`Measures{2}`	Total DTW warp cost in times
`Measures{3}`	`Measures{1} + Measures{2}`

Also note that the spectrograms are normalized ($sg = sg./max(sg(:))$) before the DTW distances are computed. This can be parameterized in a later version to enable or disable normalization.

4.1 Window size

The choice of window size dictates the frequency resolution that will be used in the metric. Small windows are unable to capture low frequency components, but can reveal nonstationary features at a high time resolution. Larger windows capture more resolution in frequency and handle lower frequency features, but lose resolution in the time domain and fail to reveal nonstationary features.

5 Notes

- The DTW algorithm that was used is **strongly assumed to be sub-optimal for this purpose**. It should be tuned to penalize sequence shifts significantly more when comparing spectra, as a shift in frequency is significantly more important than a mismatch in amplitude for a given frequency. The DTW algorithm used should be appropriate for comparing the behavior of a single frequency over time though. This will require additional work to build a more tunable DTW algorithm.

- A brief look at the Cepstrum (the Fourier transform of the log of the absolute value of the Fourier transform of the signal) didn't reveal much of interest, but this is likely due to not looking at appropriate portions of the Cepstrum. This should be looked at in the future, possibly by computing the DTW distance between relevant portions of the Cepstrum for each signal.